DOI: 10.20357/B7NG6Z

Lawrence Berkeley National Laboratory

Curation of Ground-Truth Validated Benchmarking Datasets for Fault Detection & Diagnostics Tools

Armando Casillas[1], Guanjing Lin[1],
Jessica Granderson[1], and Piljae Im[2]

[1]Lawrence Berkeley National Laboratory,
[2]Oak Ridge National Laboratory

Energy Technologies Area
August 2020

Curation of Ground-Truth Validated Benchmarking Datasets for Fault Detection & Diagnostics Tools

Armando Casillas, Guanjing Lin, and Jessica Granderson, Lawrence Berkeley National Laboratory
Piljae Im, Oak Ridge National Laboratory

ABSTRACT

Fault detection and diagnostics (FDD) analytical tools for heating, ventilation and air conditioning (HVAC) systems represent one of the most active areas of smart building technology development. A diversity of techniques is used for FDD analytics, spanning physical models, black box, and rule-based approaches, and researchers continuously strive to develop improved algorithms. With FDD algorithm numbers now in the hundreds, there is a need for performance evaluation of these algorithms in order to assess improvements, improve cost-effectiveness, and to prioritize investment in the further development of these technologies. A persistent challenge of FDD advance has been the lack of common datasets to benchmark the performance accuracy of FDD algorithms.

This paper summarizes the successful curation of HVAC operational data, paired with validated ground-truth information regarding the presence and absence of faults. The current dataset, consisting of both simulation and experimental data, will evolve to include a larger set of HVAC systems with the objective of creating the largest publicly available dataset to be used by FDD developers, users, and researchers to compare and contrast performance accuracy across FDD algorithms, helping to drive improvements that will spur greater market adoption of FDD tools. Furthermore, in order to avoid previously observed issues with contributed datasets and ensure high quality and consistency of future submissions, the development of data validation and ground-truth assessment protocol is detailed in this study.

Introduction

Corrections to existing commercial building controls infrastructure and subsequent improvements to operational efficiency can reduce building energy use by an average of 29%, which accounts for approximately 5% of overall national energy consumption (Fernandez 2017) A primary method of correcting building controls and operation is through algorithms developed to perform automated fault detection and diagnostics (FDD), which use building data to identify the presence of faults and potentially isolate root causes. As building data becomes more readily available, and as the budding field of data science and analytics comes to buildings, FDD is of increasing relevance to the research and product development communities. Outside of the research community, building owners and operators are already leading the adoption of FDD technology, using it to enable median whole-building portfolio savings of 7% (Kramer et. al. 2019).

A diversity of techniques is used for the development of FDD, spanning physical models, black box, and rule-based approaches, and developers continuously strive to develop improved algorithms (Kim and Katipamula 2018). A persistent challenge, however, has been the lack of common datasets and test methods to benchmark the performance accuracy of FDD methods, and gauge improvement of these tools over time. A few studies have been conducted to evaluate FDD algorithms and protocols (Braun, Yuill, and Cheung 2012) (Breuker and Braun 1999a) (Yuill and Braun 2013), (SCE 2015) for roof-top units. While Granderson (2018) most recently developed a test and benchmarking framework for FDD algorithm performance, demonstrating a growing need for HVAC fault datasets that can be used to further determine the accuracy and effectiveness of FDD algorithms.

This paper summarizes the successful curation of FDD test dataset with verified ground truth information on the presence and severity of faults. The data comprises simulated and experimental time series HVAC operational data (e.g. temperatures, pressures, control signals, component status, etc.) under a diversity of operating and weather conditions, combined with information on the presence and absence of faults and their associated intensity. The dataset is created with assistance from a number of contributors from the research community and spans a wide range of commercial building HVAC system configurations including: single and dual duct hydronic air handling units (AHUs), packaged roof-top units (RTUs) as well as chiller and boiler plants. The paper also details a data validation and ground truth assessment protocol for the successful development of FDD test dataset. The protocol includes a robust set of criteria that will be applied across contributors to ensure consistency of high data quality and naming schemes.

Fault Detection and Diagnostics Test Datasets

Phase 1 Preliminary Dataset

In the first phase of this study, the preliminary dataset comprises of five AHUs and one RTU, were created either through simulation, or in physical experimental facilities by multiple contributors. The dataset is stored on figshare (Granderson 2019). Table 1 summarizes the faulted and unfaulted scenarios for the preliminary dataset, including the HVAC systems, data types/facilities, fault types, and number of faulted and unfaulted days for which data were acquired for each system. For each system-fault-severity condition, a "test case" lasts for one day. More detailed information is documented in Granderson (2020).

Table 1. The systems, data types/facilities, fault types, and number of days of the preliminary dataset; SD = Single-duct, MZ = Multi-zone, VAV = Variable air volume, SZ = Singe-zone, CAV = Constant air volume, OA = Outside air, HC = heating coil, CC = cooling coil

System	**Data Type**	**Fault Types**	**Number of Days**	
			Faulted	Unfaulted
SDMZ AHU-VAV	Modelica-EnergyPlus co-simulation	OA temp sensor bias	168	28
SDMZ AHU-VAV	Iowa Energy Resource Station facility, experimental	HC valve leakage	3	7
SDMZ AHU-VAV	HVACSIM+ simulation	OA damper stuck HC valve leakage CC valve stuck	13	13
SDSZ AHU-CAV	LBNL FLEXLAB research facility, experimental	OA damper stuck HC valve stuck/leakage CC valve stuck/leakage	12	1
SDSZ AHU-VAV			7	4
RTU	ORNL 2-story flexible research platform (FRP), experimental	Condenser fouling HVAC setback error, delayed onset HVAC setback error: early termination No overnight HVAC setback Thermostat measurement bias	9	7

Phase 2 Scaled Dataset

In the second phase of the study, a scaled dataset is being created for a larger set of HVAC systems, specifically, single-duct AHU, dual-duct AHU, terminal variable air volume (VAV) boxes, terminal fan coil units, rooftop unit, chiller and boiler plants. The datasets will span a range of seasons, and operational conditions (weather, loads, etc.). Table 2 shows the HVAC systems, data types/facilities, and fault types for planned data curation. For the seven systems listed in the table, we target at least five common faults, and at least four fault severities. Data will be obtained under fault free and faulty conditions. For each system-fault-severity condition, a "test case", at least 1 day of experimental data (minute-level sampling frequency) will be collected for each season of the year. Simulation affords the opportunity to capture a full

365 days of operation for each system-fault-severity condition. The full year of simulated test cases for each system can be particularly useful to satisfy the increased training requirements of emerging learning-based FDD algorithms.

Table 2. The systems, data types/facilities, and fault types of the scaled dataset (CC - cooling coil, HC - heating coil, OA - outside air, SA - supply air)

System	**Data Type**	**Fault Types**
Single-duct Multi-zone AHU	Modelica- EnergyPlus co-simulation	CC valve stuck/leakage OA damper stuck OA/SA temp sensor bias
Dual-duct AHU Terminal VAV box Terminal fan coil	HVACSIM+ simulation	CC valve stuck/leakage HC valve stuck/leakage Cooling/heating/OA damper stuck OA/SA temp. sensor bias Cooling/heating/damper unstable control
RTU	ORNL 2-story flexible research platform (FRP), experimental Field-measured Modelica simulation	Nonstandard refrigerant charging Evaporator fouling Condenser fouling Refrigerant liquid- line restriction Presence of non- condensable in refrigerant OA damper stuck OA/SA temp. sensor bias
Chiller plant Boiler plant	Modelica- EnergyPlus co-simulation	Chiller condensers/ cooling tower / boiler fouling Chillers /pumps/ cooling towers/ boiler unstable control The condenser water leaving the three-way valve stuck/leakage Temperature sensor bias Pressure sensor bias

FDD Test Datasets Documentation

The data set is documented in a common format. The documentation of the datasets is broken into two parts: a text document (pdf) containing the ‘metadata' information and comma-separated values (CSV) files storing the time series data. The ‘metadata’ information can help the data set users map the data points to required algorithm inputs and configure algorithm thresholds. The trended data in the CSV files are the actual data inputs for the targeted FDD tools.

Each CSV file represents a single combination of system configuration and experimental or simulated data creation approach. The data are minute-frequency time series measurements of the system operational parameters that are most commonly available to FDD algorithms in typical commercial buildings. Time stamps are in the first column of each file, and presented in the format m/d/yy h:mm. The final column of each file contains a binary indicator of the ground truth information on whether or not a fault is present.

In the text document of 'metadata' information, each dataset is documented according to a common template to present the key information necessary to understand the content and scope of the dataset. The template includes four sections: "Dataset Overview", "Building and System Information", "Data Collection", and "Input Scenarios (faulted or unfaulted)".

The Data Overview Section gives an overview of the dataset, including who created it, and whether it is generated through simulation or physical experimentation. For example, the experimental fault dataset for RTU-VAV system was generated by Oak Ridge National Laboratory (ORNL) in ORNL's light commercial 2-story flexible research platform (FRP).

The Building and System Information Section describes the simulation model or experimental facility, illustrates the type and configuration of the system with the schematic diagram, and highlights the control sequence. The schematic diagram shows the main components of the system and the sensors installed. The control sequence summarizes the operation modes (e.g. occupied, unoccupied modes) and the control strategies as well as setpoints under each operation mode. Figure 1 shows a schematic diagram example for a RTU experimental fault dataset. This is a Trane® YCD150 12.5-ton rooftop unit. It has two-stage compressors, cooling coil, heating coil, outside and return air dampers, and supply air fan. The unit serves a total of 10 zones in FRP. The control sequence used by this unit is summarized below:

"The RTU is scheduled for automatic operation on a time of day basis for occupied and unoccupied mode. The occupied mode starts at 7:00am and ends at 10:00pm .
Occupied mode

- *Supply air temperature (SAT) control: two compressors and gas furnace shall on/off to maintain a SAT setpoint. The SAT setpoint is 55°F year-round.*
- *Space temperature control: The zone heating and cooling setpoint are 69.8°F and 75.2°F during the occupied time period.*

Unoccupied mode

- *Unoccupied heating: zone air temperature heating setpoint is 60°F.*
- *Unoccupied cooling: zone air temperature cooling setpoint is 80°F."*

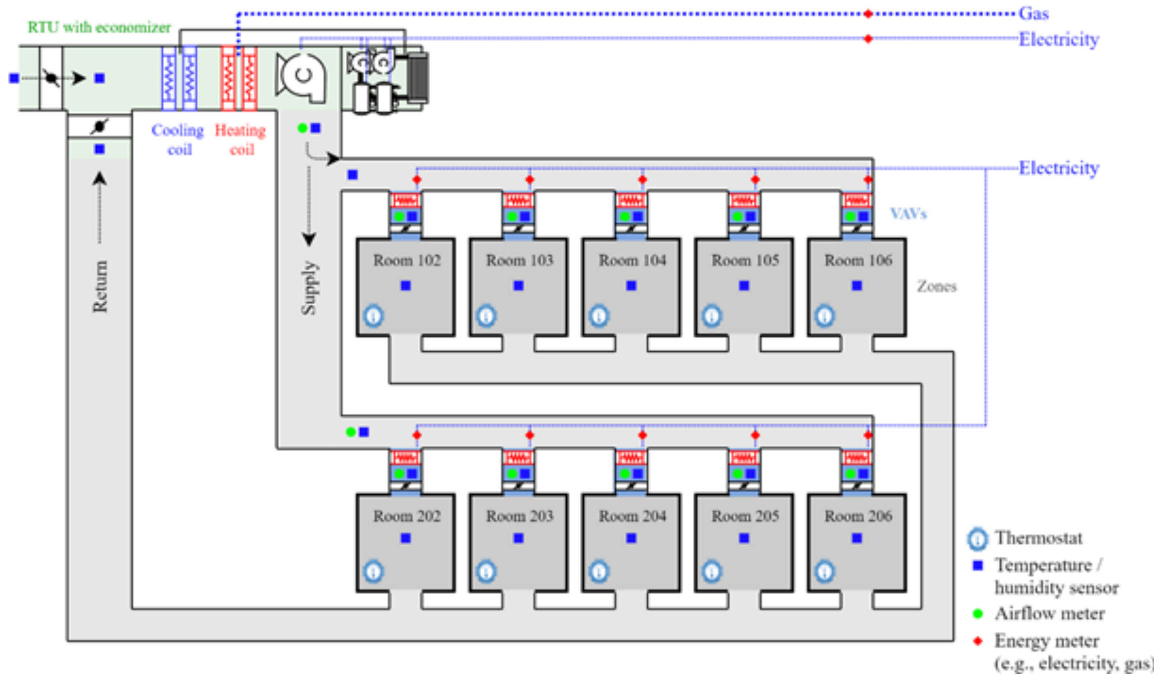

Figure 1. Example schematic diagram of RTU and the connected ten VAV boxes that serve ten zones

The Data Collection Section summarizes the name, description, unit, and time interval of the data points being collected in the simulation or experiments. Table 3 shows the data points covered in an experimental RTU data set. The data points are recorded at 1-min intervals. The Input Scenarios Section describes all the faulted and fault-free conditions represented in the data. The fault types, intensities, method of fault imposition, and fault occurred time are summarized in the table. For instance, Table 4 shows the three levels of intensity of nonstandard refrigerant charging fault was imposed into the RTU during the month of August in 2018. Each fault-intensity scenario lasted for one day.

Table 3. Example data points for the experimental RTU data

Data Point Name	**Description**	**Unit**
RTU: Outside Air Temperature	Measured RTU outside air temperature	°F
RTU: Supply Air Temperature	Measured RTU supply air temperature	°F
RTU: Return Air Temperature	Measured RTU return air temperature	°F
RTU: Supply Air Fan Status	RTU supply air fan status; 0-off, 1-on	--
RTU: Circuit Discharge Temperature	Array of refrigerant temperature on the RTU	°F
RTU: Circuit Condenser Outlet Temperature		
RTU: Circuit Suction Temperature		
RTU: Circuit Discharge Pressure	Array of refrigerant pressures on the RTU	PSIA
RTU: Circuit Condenser Outlet Pressure		
RTU: Circuit Suction Pressure		
RTU: Supply Air Volumetric Flow Rate	Measured RTU volumetric air flow rate	Cfm
RTU: Electricity	RTU electricity consumption	Wh
RTU: Fan Electricity	RTU supply air fan electricity consumption	Wh
RTU: Natural Gas	RTU natural gas consumption	Cfm
Occupancy Mode Indicator	Indicator if the system operates in occupied mode; 1-occupied mode, 0-unoccupied mode	--
RTU: Compressor On/Off Status	RTU compressor status; 0 – off, 1 - on	--
Fault Detection Ground Truth	Indicator if there is a fault present; 0 – unfaulted, 1 - faulted	--

Table 4. Example input scenario included in experimental RTU fault datasets

Input Scenarios		Method of fault imposition	Fault occurred time
Fault type	**Fault intensity**		
Nonstandard refrigerant charging	-15% undercharged	Introduce low, or excessive mass of refrigerant in line	8/14/18
	-30% undercharged		8/16/18
	15% overcharged		8/18/18-8/20/18

Data Validation, Ground Truth Assessment Protocol

In order to ensure consistent, clean, and accurate fault datasets that can then be applied to evaluate 3rd party FDD algorithm performances with confidence, a predetermined criterium has been developed and detailed in this section. It is important to be able to get contributions from a wide variety of research groups, system expertise, modeling capabilities, and test facilities. To successfully synthesize all of the contributed data, and to avoid previous challenges and difficulties with contributed datasets, some level of standardization and common format is required. We present some of the conventions established for this work - other choices are possible, but this is what has been defined for the fault data library. This should be true of each contributed dataset in terms of quality assurance for the accuracy of ground truth and fault symptoms. Early experience working across multiple data contributors has surfaced the need to consistently ensure data quality and ground truth (fault presence or absence) accuracy. Below are examples of quality assurance considerations deemed necessary for both experimental and simulated cases. Examples of important quality assurance checks for contributors to make are:

- Complete documentation on system control sequences, complete list of relevant measurements for fault evaluation
- Organized and well formatted datafiles, consistent naming scheme for points
- Presence of all required, essential measurement points
- Realistic measurements (for simulated data) or, clean and calibrated field and sensor data

In order to effectively develop a high-quality fault database for a variety of systems, a protocol was established to validate that all datasets meet predetermined criteria. Below is a general check list that would be applied to all parties interested in submitting a data set to the library. Additionally, these criteria will be denoted by the following letters, which signify the phase in which these aspects should be checked throughout the data creation process.

- Initial check **(I)**: Before creating data
- Very early check **(E)**: Before scaling data
- Verification at scale **(S)**: When high volume of data set is available

System info, control sequence documentation (I, E):

A detailed document with information regarding the system configuration must be available to determine expected behavior of a fault-free system. System information also includes any diagrams or as built drawings that may exist. Submitters must provide appropriate nameplate information for all systems (AHU's, Fans, RTU's). Nameplate information is to ensure information such as data measurement limits and expected fault free behavior is established. All relevant control sequences for systems pertaining to the dataset must have documentation. For an economizing multi-zone VAV HVAC system for example:

- **Occupancy/Schedules**: occupied/non occupied days and time windows, equipment and zone schedules
- **Temperature Control**: supply air, zone temperature heating and cooling setpoints
- **Humidity Control:** supply air, zone relative humidity setpoints
- **Pressure control:** supply air static pressure setpoints
- **Actuating Component Control:** damper control sequences (outdoor air, terminal VAV dampers) as well cooling and heating components (cooling and heating/reheat valves)
- **Motor Component Control**: Fan speed (for variable frequency drives), and minimum compressor run-time (for RTU's)

Contributors should state if the system is controlled by standard industry protocols (e.g ASHRAE Guideline 36).

Sensor accuracy and calibration, functional testing (I):

Experimental facilities must ensure that all sensors, meters and actuators have been recently calibrated and full functional testing has been conducted on all components of their system including damper and valve positions.

Diversity of weather, operational conditions (I, E, S):

To properly assess the change in FDD performance over different ambient conditions, we need a diversity of operational conditions for each fault-severity combination. For experimental test fault data, it is preferred to have at least 3 days worth of data pertaining to each fault condition and severity for at least 3 seasonal conditions (summer, winter, shoulder). Simulated data is required to contain at least 1 week of fault conditions at each severity for all 4 seasons (spring, summer, fall, winter). This ensures that all fault symptoms are observed throughout varying weather conditions, with the expectations that some faults may show more severe symptoms during certain seasons. This variation in symptom severity can then be applied to a FDD tool's ability to detect these faults throughout the year. In addition to symptom severity, operational conditions change throughout the year, with heating control mostly present in the winter months, while economizing sequences mostly operate under mild weather conditions during shoulder seasons; therefore, diversity of seasonal data allows for a more holistic view of system performance.

Data quality (E,S):

Perhaps most importantly, contributing datasets must undergo a quality check protocol and meet the following criteria to be considered suitable for inclusion in the fault library. All data must conform to this common convention so that a FDD algorithm/tool can streamline analysis of the full dataset as a batch, without lots of tweaking by the tester.

Data set limit:

Each sensor data stream must have its data fall within physical limits of the system it is measuring, an example of these limits are seen below for the experimental facility's 12.5 ton RTU-based HVAC system:

Refrigerant Temperature: -100F - 400F
Refrigerant Pressure: 0 - 600psia
Air Temperature: 0 - 125 F
Air RH: 0 - 100 %
Airflow (depends on size of system): 0 - 50000cfm
Electricity Consumption: 0 - 1000kWh
Natural Gas Consumption: 0 - 10cfm
Status, Commands, Control Signals ,GT (binary or fractional): 0-1

Static Reporting Frequency, Timestamp:

Each datafile must be reported in timeseries format at a consistent reporting frequency (e.g., 1 minute, or 15 minute). Additionally, timestamps should be listed as the first column in the datafile.

Missing data:

Although it is preferred that there be no gaps in the timeseries for each data point, it is acknowledged that this is difficult in field test conditions. Each measurement data stream is allowed to have at most 20% of its data missing in total and at most 10% during occupied times.

Datafile type, Data Completeness and Format: (I):

All files must be submitted in CSV form, or similar delimited file with headers and timeseries column. All submitted data files should also follow the same naming schema and format, for example:

[System_#]:[Component]+[Measurement]

e.g., AHU_1: Supply Air Temperature
AHU_2: Cooling Coil Valve Control Signal

Additionally, all datasets must include all relevant measurements related to the fault. A high volume of data will be generated under different operation conditions. However, those variables can cause a "data rich, but information poor" status, i.e., not all variables are equally important to evaluate the equipment operation and needs to be used to validate the simulation results. Therefore, it is necessary to first determine the key measurement which can be used as indicators to reflect equipment/system operation performance. Those key measurements can also be employed to validation the faulty data set, as well as fault-free data set. Meaning if said fault is an air-side fault, all measurements pertaining to air conditions (temperature, RH, pressure, volumetric flow) and all components potentially affecting the air conditions (fans, dampers, coil valves) must be present.

Check against fault free operation (E,S):

Under fault free conditions, we will need to verify whether the system operates under the designed control sequence and that system operates within or at setpoint, and overall reflective of fault-free behavior. A failed fault free check would indicate an incorrect implementation of a given control, as well as equipment faults, or inaccurate reporting of status and command data for system components. For simulations under fault-free conditions, steady state and dynamic experimental data may be used to compare to expected modeled system behavior.

Check against expected fault symptoms with measured values (E,S):

Data that is reflective of fault-present operations is verified by comparing symptoms observed during testing with expected fault symptoms. Fault symptoms are defined through a literature review process, which allows measured test values to be compared to expected trends for faulty cases. All submitted faults must have proof of ground truth fault conditions and must reflect expected behavior for said fault. The first type of fault symptom is the changes in the value of a given measurement. For example, the changes in the value of a given measurement of a non-standard refrigerant charging(undercharging) is provided in Table 5. The plus marks indicate values that are expected to increase, while values with a minus symbol have been known to decrease with the fault in question present, lastly a 0 indicates a value with little or no change.

In an RTU vapor compression system, lack of refrigerant mass in the refrigerant line (pounds of refrigerant) leads to modification of heat transfer through its evaporator and condenser due to less working fluid being present in each component at any given time. Undercharging results in increased temperature difference evaporator, higher evaporator and discharge superheat and COP degradation and increased overall electric consumption. Low refrigerant charge levels will also lead to starving and short cycling of the compressor due to overheating and eventually potential component failure (Heinemeier 2012) (Breuker and Braun 1999a, 199b) (Yuill and Cheung 2013) (Shen, Braun, and Groll 2009) (SCE 2015)

Table 5. Fault Symptoms for Nonstandard Refrigerant Charging (Under Charging) based on Literature

Measurement Type	**Measurement Point**		
Temperature (F)	**RTU: Discharge Temperature**	**RTU: Condenser Outlet Temperature**	**RTU: Suction Temperature**
	+	+	+
Pressure (PSIA)	**RTU: Discharge Pressure**	**RTU: Condenser Outlet Pressure**	**RTU: Suction Pressure**
	-	**0**	**-**
Electric Energy (kWh)	**RTU: Electricity**		
	+		

Checking against expected faulted symptoms with the RCA Protocol (S):

Fault symptoms are also checked with the changes of physical relationships between multiple measurements. For instance, the RCA Protocol, a known FDD protocol found in the appendices of California's Title 24 building code (CEC 2019) can be used for symptom checking of a nonstandard refrigerant charging fault and evaporator airflow fault data set. Figure 2 shows the flowchart of the RCA protocol. The RCA protocol uses supply and outdoor air dry bulb temperature, return air wet bulb temperature and refrigerant subcool and superheat temperatures to determine evaporator airflow reduction and nonstandard refrigerant charging faults.

When the unit is under steady-state operation, the system conditions must fall under a certain range of temperatures for dry and wet bulb return air temperatures. If the difference between the target temperature split and the temperature difference between measured drybulb supply and return air temperature is larger than 3 °F, an evaporator air flow fault is concluded. When the split is below 3 °F, the logic then checks for a possible refrigerant charge fault. Given whether the system has a fixed orifice (FXO) metering device or a thermal expansion valve (TXV) the measured subcool or superheat temperatures of the refrigerant are compared to a target temperature based on drybulb ambient temperature and wetbulb return air temperature. An overcharging fault is diagnosed if the difference in superheat or subcool temperatures are greater than 5 °F or 3 °F respectively, and an undercharging fault is identified if the split is less than -5 °F or -3 °F respectively. Figure 3 shows an example of the checking results for a RTU experimental refrigerant undercharge fault data set. During the majority of the time the compress is on (compressor status = 1), the undercharge fault is diagnosed (fault status = -1) with the RCA protocol.

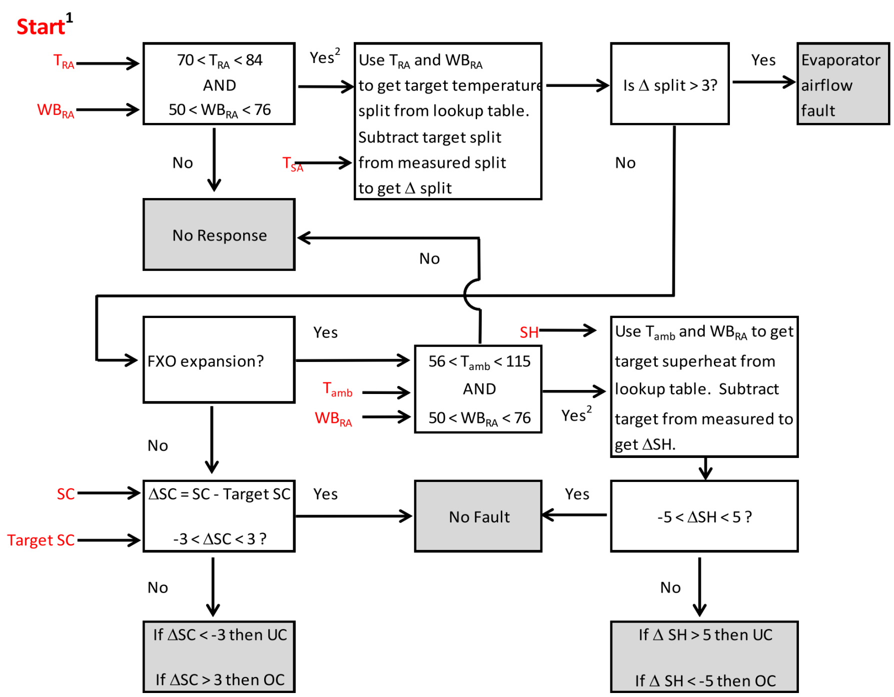

Figure 2. Logical flow chart for CEC's RCA Protocol, a diagnostic method for evaporator and refrigerant charging faults in packaged RTU systems. (T_{RA} – dry bulb return air temperature, WB_{RA} – wet bulb return air temperature, T_{SA} – dry bulb supply air temperature, FXO – fixed orifice, T_{amb} – dry bulb ambient temperature, SH – refrigerant superheat temperature, SC – refrigerant subcool temperature)

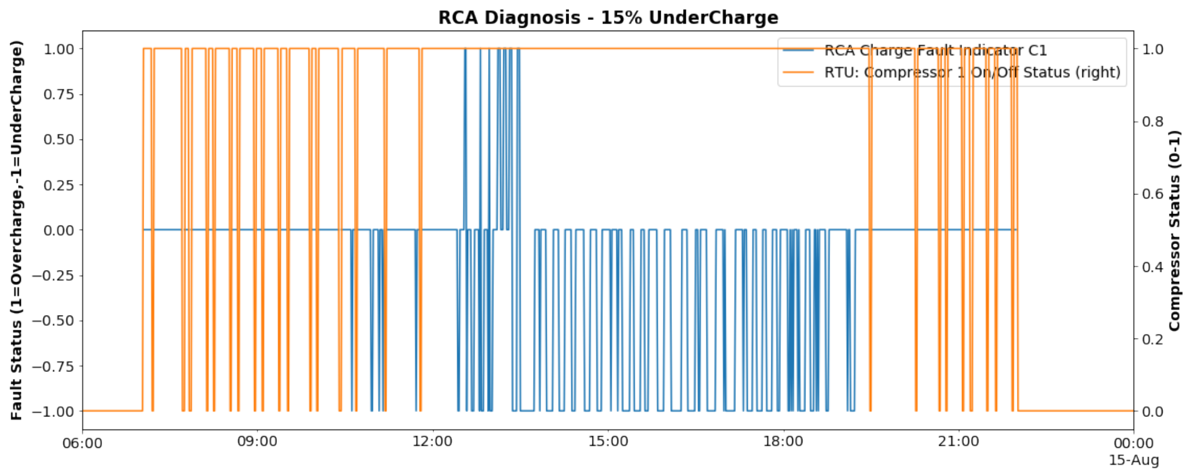

Figure 3. Example of RCA Protocol diagnosis on an experimental 15% undercharge fault dataset, the fault is shown to be reading a true positive for the majority of the time the compressor is on

Conclusion and Future Work

With the upsurge in software, data availability, and data analytics across the buildings industry, new FDD algorithms are continuously being developed. There is a lack of standard datasets for evaluating the accuracy of FDD algorithms and the improvement overtime. In response, this paper summarizes the curation of FDD test dataset which are the HVAC operational data, paired with validated ground-truth information regarding the presence and absence of faults. The datasets are created by multiple contributors and synthesized into a single repository with a common format and documentation. The paper also illustrates a data validation and ground-truth assessment protocol which are a set of predetermined criteria to ensure collect consistent, clean, and accurate datasets from the contributors. The validation of a fault dataset is then followed by applying the dataset to a FDD protocol to demonstrate a possible use case of these data. The performance of this protocol, along with others, can be further evaluated with the expansion of this fault library.

In the future, we will continue the creation of the scaled FDD test dataset with the application of the ground truth data validation protocol, and provision of the dataset for public use by the FDD research and development community.

Acknowledgement

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We also recognize each of the fault detection and diagnostic tool developers who participated in this survey. We would also like to thank Erika Gupta and the Building Technologies Office as well as our data contributors.

References

Braun, J.E., Yuill, D. , Cheung, H. (2012). Final Project Report a Method for Evaluating Diagnostic Protocols for Packaged Air Conditioning. *California Energy Commission*.

Breuker, M. S., and Braun, J. E. (1999a). Evaluating the performance of a fault detection and diagnostic system for vapor compression equipment. *ASHRAE Transactions*, *105*.

Breuker, M. S., and Braun, J. E. (1999b). Common faults and their impacts for rooftop air conditioners. *ASHRAE Transactions*, *105*(April 2013), 37–41.

Cheung, H., and Braun, J. E. (2013). Simulation of fault impacts for vapor compression systems by inverse modeling. Part II: System modeling and validation. *HVAC and R Research*, *19*(7), 907–921. https://doi.org/10.1080/10789669.2013.819769

California Energy Commission. (2019). *2019 Title 24: BUILDING ENERGY EFFICIENCY STANDARDS*. (June 2019).

Fernandez, N, et al. 2017. Impacts on commercial building controls on energy savings and peak load reduction. Pacific Northwest National Laboratory. PNNL Report Number PNNL-25985

Granderson, J., Lin, G., Harding, A., Im, P., Chen, Y. Dataset for building fault detection and diagnostics algorithm creation and performance testing. figshare https://figshare.com/s/c0aadb2bfc76ea3c3424 (2019).

Granderson, J., Lin, G., Harding, A., Im, P., and Chen, Y. (2020). Building fault detection data to aid diagnostic algorithm creation and performance testing. *Scientific Data*, *7*(1), 1–14. https://doi.org/10.1038/s41597-020-0398-6

Granderson, J., Lin, G., Singla, R., Mayhorn, E., Ehrlich, P., and Vrabie, D. (2018). Commercial Fault Detection and Diagnostics Tools: What They Offer, How They Differ, and What's Still Needed. *Lawrence Berkeley National Laboratory Recent Work*, 1–12. https://doi.org/10.20357/B7V88H

Kim, W., and Katipamula, S. (2018). A review of fault detection and diagnostics methods for building systems. Science and Technology for the Built Environment, 24(1), 3–21. https://doi.org/10.1080/23744731.2017.1318008

Kramer, H, Lin, G, Curtin, C, Crowe, E, Granderson, J. (2019) Building Analytics and Monitoring-based Commissioning: Industry Practice, Costs, and Savings. In Press, Energy Efficiency

Lin, G., Kramer, H., and Granderson, J. (2020). Building fault detection and diagnostics: Achieved savings, and methods to evaluate algorithm performance. *Building and Environment*, *168*(January). https://doi.org/10.1016/j.buildenv.2019.106505

Shen, B., Braun, J. E., and Groll, E. A. (2009). Improved methodologies for simulating unitary air conditioners at off-design conditions. *International Journal of Refrigeration*, *32*(7), 1837–1849. https://doi.org/10.1016/j.ijrefrig.2009.06.009

Yuill, D. P., and Braun, J. E. (2013). Evaluating the performance of fault detection and diagnostics protocols applied to air-cooled unitary air-conditioning equipment. *HVAC and R Research*, Vol. 19, pp. 882–891. https://doi.org/10.1080/10789669.2013.808135

Southern California Edison. (2013). *Evaluating the Effects of Common Faults on a Commercial Packaged Rooftop Unit*. (July).

Southern California Edison. (2015). *Laboratory Assessment of Fault Detection and Diagnostics Technologies on a Commercial Packaged Rooftop Unit*. (July).